BACTERIOLOGICAL REVIEWS, Dec. 1967, p. 215-229
Vol. 31, No. 4

Printed in U.S.A.

Arrangement of Base Sequences in Deoxyribonucleic Acid[1]

BRIAN J. McCARTHY

Departments of Microbiology and Genetics, University of Washington, Seattle, Washington 98105

INTRODUCTION

There are several levels at which the arrangement of nucleotides in deoxyribonucleic acid (DNA) may be discussed. At the gross level, a considerable body of information exists concerning the overall base composition of DNA. This has proven valuable for taxonomic purposes, at least in microorganisms, and provides some comparative information as to the relative variability of the composition of DNA among various groups of organisms. In higher plants and metazoans, the base composition is relatively constant, so that these measurements are not sensitive enough for comparative purposes. At a higher level of resolution, nearest-neighbor relationships may be determined with the aid of DNA polymerase (21), and this may be extended to larger nucleotide sequences by allowing the incorporation of a mixture of ribonucleotides and deoxynucleotides into the product in the presence of Mn^{++} (4). In the case of nearest-neighbor frequencies, 16 parameters are provided for the characterization of a given DNA. The method has been used for evolutionary studies with viral DNA (2, 37), although it is clearly limited in sensitivity.

The ultimate description of a DNA molecule or of a gene is in its complete base sequence. Complete nucleotide sequence data are available only for a few soluble ribonucleic acid (sRNA) molecules (18, 41, 57). There is no general method for the isolation of a single gene, except through the purification of a single messenger which in turn is possible only in simple viral systems (1). When this is achieved, sequence determination is much more difficult than amino acid sequence determination in proteins, for a number of obvious reasons. There are three times the number of residues in the messenger ribonucleic acid (mRNA) or a single strand of the DNA as compared to the gene product, and only four kinds of bases compared to 20 kinds of amino acids. The existence of minor bases in sRNA has been as valuable to the determination of sequence as the small number of residues. Again, highly specific enzyme degradation methods other than limited digestion are not available.

At the moment, therefore, the problem of complete sequence determination of even the smallest viral nucleic acids is far from solution. Fortunately, however, the great power of nucleotide sequence recognition which nucleic acids themselves possess may be used as a tool for detailed comparison of related DNA and ribonucleic acid (RNA) molecules. After the discovery by Doty et al. (12) and Marmur and Lane (24) that the two strands of DNA may be separated and reannealed, it became clear that this reaction could be extremely valuable for comparisons of related nucleic acids as well as the analysis of RNA produced from the DNA. Thus, the formation of heteroduplex DNA molecules by the separated DNA strands originating from two organisms has become an accepted method in taxonomy (19, 26, 31). This method is appropriate with any group of DNA molecules from bacteria to primates. An example of these results is shown in Table 1, in which various enterobac-

[1] Eli Lilly Award Address (1967).

TABLE 1. *Reaction among heterologous DNA preparations*

DNA prepn	Reaction
	%
Bacterial[a]	
Escherichia coli	100
Salmonella typhimurium	71
Aerobacter aerogenes	51
Proteus vulgaris	14
Serratia marcescens	7
Pseudomonas aeruginosa	1
Bacillus subtilis	1
Mammalian[b]	
Rhesus monkey	100
Human	76
Chimpanzee	76
Owl monkey	68
Potto	50
Mouse	22
Chicken	11
Salmon	5
E. coli	0

[a] From McCarthy and Bolton (31). Reactions expressed relative to the homologous reaction by *E. coli* DNA.

[b] From Hoyer et al. (20). Reactions expressed relative to the homologous reaction by rhesus monkey DNA.

terial DNA preparations are ordered according to the similarity of their base sequence with that of *Escherichia coli* DNA (31), and primate DNA preparations are compared with those of the rhesus monkey (20).

The ensuing paragraphs will be concerned not so much with these applications of the techniques, but rather with the arrangement of different nucleotide sequences in a genome. As will be clear later, the reaction of two strands of homologous DNA can provide useful information as to the frequency of occurrence of a given sequence in the whole genome and the incidence of similar, but not identical, sequences. The formation of structures other than true renatured DNA is a function of the kinds of related sequences which exist in that particular DNA.

STRAND ASSOCIATION BY DNA

General Consideration

The reassociation of two DNA strands to form renatured DNA may be studied by a variety of methods. These include changes in hypochromicity and other optical parameters, buoyant density, and biological activity (26). None of these is suited to studies of the interaction of two strands originating from two organisms or of interactions which form products other than renatured DNA. The density gradient technique may be modified to study interspecific DNA reactions by labeling one DNA with a heavy isotope (12). However, more convenient means of studying all of these various types of reassociation reactions are provided by methods in which a radio-labeled DNA is allowed to interact with a homologous or heterologous DNA immobilized in a solid phase. The latest and most convenient version of this approach is offered by Denhardt's (10) modification of the membrane filter technique of Gillespie and Spiegelman (15), originally devised for assay of DNA-RNA hybrids.

At reasonably low concentrations of DNA (<10 μg/ml), renaturation of DNA in solution follows second-order kinetics, as would be expected from the interaction of separated DNA strands (49). At higher concentrations or in high ionic strength, aggregation of renatured DNA complicates the kinetics (48, 49). The rate of renaturation is clearly dependent upon the number of different base sequences in the DNA. Thus, DNA representing a simple genome, such as that of a virus, renatures more rapidly than that of a more complex bacterial or mammalian genome when incubated at the same concentration in accordance with the higher concentration of particular complementary elements (6, 25). Under the nonideal, but more convenient, conditions of reaction of labeled DNA in solution with unlabeled filter-bound DNA, the kinetics are more complex. As shown in Fig. 1, the rate of reaction is proportional to the concentration of radiolabeled DNA in solution. With small amounts of labeled DNA in solution where the reaction between two labeled DNA strands is unimportant, the reaction rate is proportional to the amount of DNA adsorbed on the filter over a limited range. With larger amounts of filter-bound DNA, the reaction rate is not limited by the quantity of immobilized DNA (Fig. 2). These facts complicate the interpretation of reaction rate data, although, as will be discussed later, the rate of reaction does depend on the relative homogeneity of the DNA sample in question.

Two important kinds of measurement applied to various types of DNA will be discussed. The rate of reassociation of two strands at various temperatures is a measure of the number of elements or sequences in that particular DNA which are similar enough to form a duplex structure. When true renaturation is being measured, the rate of reaction is proportional to the complexity of the DNA or the genome size measured in nucleotide residues. Under less stringent conditions, such as lower temperatures, the rate depends upon the incidence of similar,

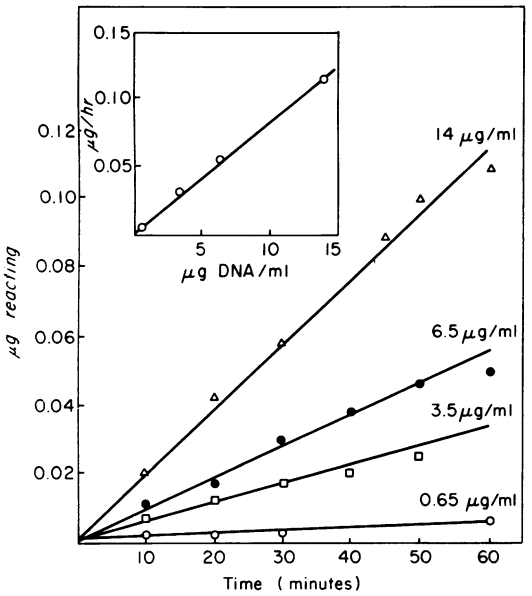

FIG. 1. *Dependence of the rate of reaction of sheared denatured DNA with filter-bound denatured DNA on the concentration of DNA in solution. Various quantities of ^{32}P-labeled sheared T4 DNA were incubated at 60 C with a 5-mm filter containing 12 μg of denatured T4 DNA in 0.2 ml of 1 × SSC (see Fig. 3). The amount of complex formed was assayed by counting filters removed at various times.*

but not identical, sequences distributed throughout the genome.

In addition, the properties of the complex formed under various conditions may be assessed by means of its thermal stability. As in the case of denaturation of DNA in solution, the temperature at which the duplex dissociates is a function of its base composition and the degree of base pairing which exists. The thermal dissociation of duplexes formed by incubating labeled denatured *Bacillus subtilis* DNA with homologous filter-bound DNA is illustrated in Fig. 3. The temperature of dissociation depends on the ionic strength of the solution. In the following pages, examples of these two types of measurement will be given with simple and complex DNA. The guanosine plus cytosine (GC) percentage values of the three DNA preparations to be discussed most fully are quite similar, so that differences are ascribable to complexity rather than to base composition. Values usually given for the three are: T4 bacteriophage DNA, 36% GC; *B. subtilis* DNA, 42%; mouse DNA, 41%.

Viral DNA

Under conditions of low concentration and moderate ionic strength, T4 DNA renatures in solution at the optimum temperature, 56 C, to give structures indistinguishable from native DNA (48, 49). This process has a broad optimum temperature and does not take place above 75 C or below 40 C. Similar effects can be seen in the reaction of labeled DNA with filter-bound DNA. The rate of reaction shows an optimum at about 60 to 65 C and falls off at lower or higher temperatures (Fig. 4). At about 40 C, there is a suggestion of a second component, although, as shown by the rate of reaction with heterologous DNA, this is not species-specific. In fact, the rate of reaction at 30 C is faster with *B. subtilis* or mouse DNA than with homologous DNA, since their greater complexity permits an increased possibility of complementary pair formation over short regions.

A characterization of the reaction products is illustrated in Fig. 5. At 40 C or above, the thermal stability of the products formed are identical, having a melting temperature (T_m) of approximately 82 C. This is close to that expected on the basis of the T_m for native DNA, 85 C, and a decrease of about 3 C resulting from the lower molecular weight of the sheared DNA (6). When incubated at 25 C, only complexes of low stability are formed, and these are also formed with

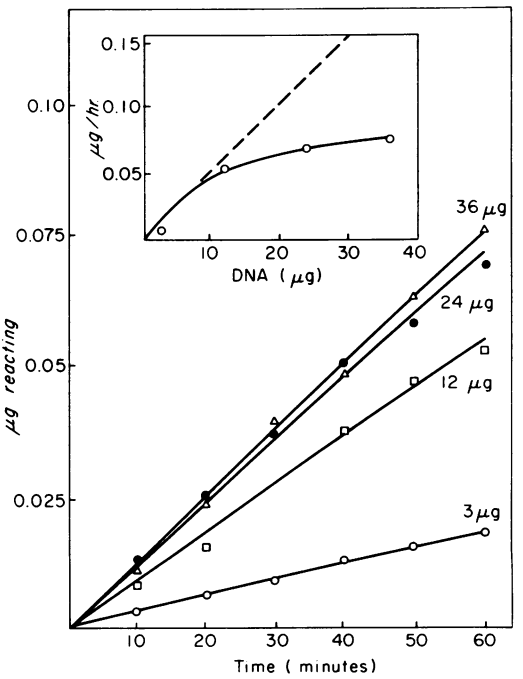

FIG. 2. *Dependence of the rate of reaction of sheared, denatured DNA with filter-bound denatured DNA on the amount of filter-bound DNA. ^{32}P-labeled, sheared T4 DNA (1.2 μg) was incubated with various quantities of filter-bound DNA at 60 C in 0.2 ml of 1 × SSC.*

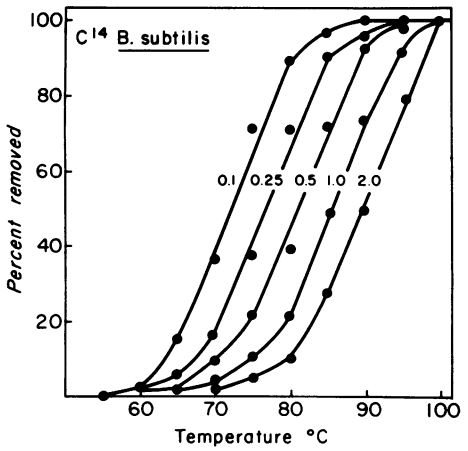

FIG. 3. *Thermal denaturation profiles of duplex formed by incubating 1 μg of ^{14}C-labeled, sheared, and denatured Bacillus subtilis DNA with filters (5 mm in diameter) containing 10 μg of denatured B. subtilis DNA. Incubations were carried out in 0.20 ml of 2 × SSC at 67 C for 15 hr. The duplex was dissociated by heating the filters in 2-ml volumes of solutions of saline-citrate of the indicated strength to various temperatures. The radioactive DNA removed at each temperature was collected by precipitation with trichloroacetic acid and counted. The numbers on each line indicate the salt concentrations in terms of standard saline-citrate (SSC), i.e., 0.15 M NaCl, 0.015 M sodium citrate. Thus, 0.1 indicates a solution of one-tenth these concentrations.*

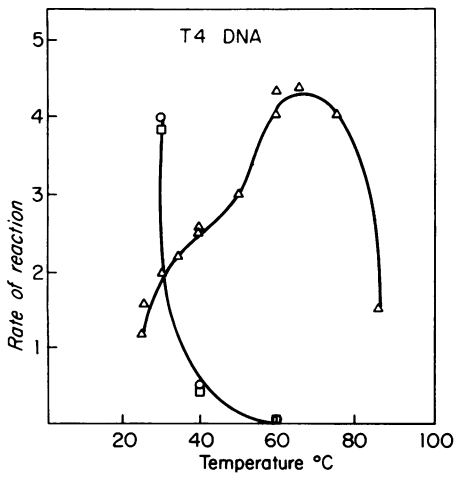

FIG. 4. *Rate of DNA duplex formation by sheared, denatured DNA of phage T4 with homologous and heterologous DNA at various temperatures. Denatured DNA (1.2 μg) was incubated with 12 μg of filter-bound DNA. Initial reaction rates are expressed as percentages of input labeled DNA reacting per hour.* (△) *T4 DNA.* (○) *Bacillus subtilis DNA.* (□) *Mouse DNA.*

heterologous DNA. There is no evidence for complexes of intermediate stability which would result from partial homology of regions in different parts of the phage chromosome. In fact, it appears that no two regions of the phage genome are any more similar in base sequence than would be expected on a random basis. Associations between these give structures with a T_m of about 35 C (Fig. 5).

Bacterial DNA

Experiments similar to those described in the previous section but with *B. subtilis* DNA give somewhat different results (Fig. 6). In the higher temperature range, reaction of labeled DNA with the filter-bound DNA shows an optimum at the

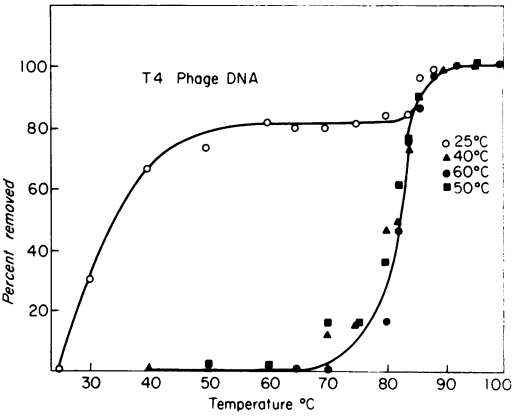

FIG. 5. *Thermal dissociation profiles of duplexes formed by incubating 1.2 μg of ^{32}P-labeled T4 DNA with 12 μg of filter-bound homologous DNA at various temperatures.*

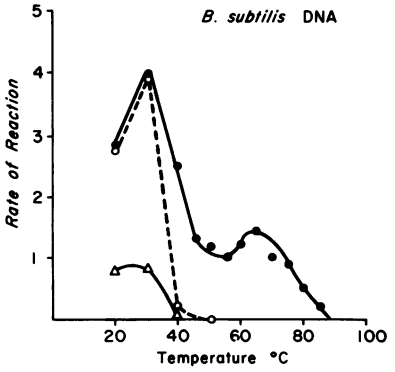

FIG. 6. *Rate of DNA duplex formation by labeled Bacillus subtilis DNA with homologous and heterologous DNA at various temperatures. Conditions as in Fig. 2.* (●) *B. subtilis DNA* (○) *Mouse DNA.* (△) *T4 DNA.*

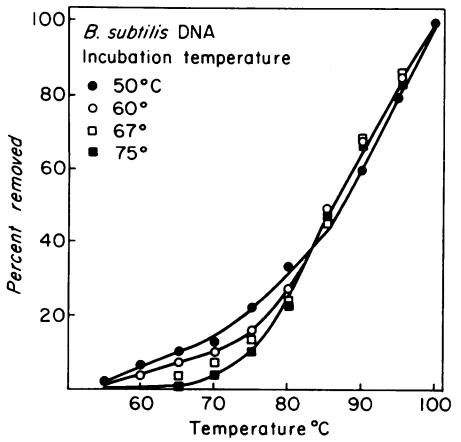

FIG. 7. *Thermal denaturation profiles of duplexes formed by incubating labeled Bacillus subtilis DNA with filter-bound B. subtilis DNA at various temperature conditions (as in Fig. 3).*

expected temperature (25). As the temperature is reduced below 45 C, however, the rate of reaction increases to a second optimum at 30 to 35 C. By comparison with reactions of labeled *B. subtilis* DNA with heterologous DNA, it appears that this component is at least partially species-specific. Thus, the reaction at 40 C with homologous DNA is at least 10 times more rapid than that with mouse or phage T4 DNA. At even lower temperatures, the heterologous reaction becomes as rapid as the homologous reaction.

The thermal stability of the various reaction products formed is shown in Fig. 7 and 8. After incubation at 50, 60, 67, or 75 C, most of the reaction product has the high stability expected of well-renatured DNA, i.e., with a T_m similar to that of sheared native DNA (Fig. 7). With incubation at 40 C, some of the product has this high stability (Fig. 8), but most is of much lower stability. With incubation at 25 C, the reaction is not species-specific and the stability is low. Thus, there is some evidence for a species-specific recognition of a very distant intragenome homology unlike that in T4 phage DNA. However, the stability is quite low and the degree of base sequence homology existing among different parts of the bacterial chromosome must be quite limited.

Mammalian DNA

It is not possible to demonstrate the renaturation of mammalian DNA by the CsCl method. On the other hand, specific strand association may be demonstrated by the DNA-agar (19) or membrane-filter method. It has been clear for some time, however, that these strand associations, athough species-specific, do not represent the formation of genuine renatured DNA. This is evident from considerations both of the rates of such reactions and of the properties of the complexes formed. These facts are illustrated by the data in Fig. 9 and 10. As is shown in Fig. 9, the rate of reaction at the optimum temperature is closely similar to that for *B. subtilis* DNA. In view of the fact (29) that the DNA content per haploid genome is some 500 × higher than for *B. subtilis* (Table 2), this reaction rate is much greater than theoretical and cannot represent the association of two complementary strands originating from the same genetic site. Either considerable duplication of sequences exists or the reactions are not completely specific. Other differences between *B. subtilis* and mouse DNA occur in the dependence of reaction rates on temperature. Although the base compositions are almost identical, the optimum temperature for reaction is lower for mouse DNA. In addition, the rate falls to extremely low levels above 70 C, whereas the rate of reactions of *B. subtilis* DNA falls only slowly.

The thermal dissociation profiles for the complexes formed are highly dependent upon the temperature of incubation (Fig. 10). Higher incubation temperatures give rise to complexes of higher stability. Thus, a great variety of species-specific complexes may be formed whose relative

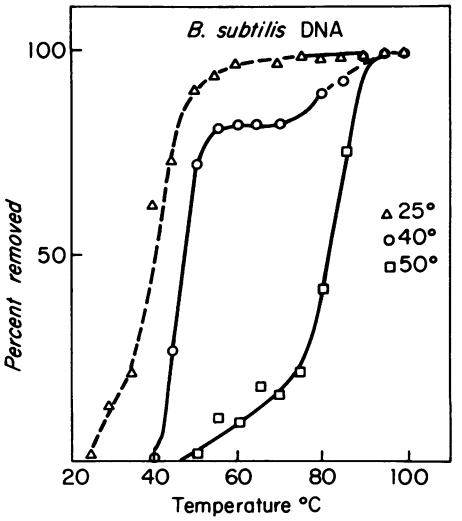

FIG. 8. *Thermal dissociation profiles of duplexes formed by incubating labeled Bacillus subtilis DNA with filter-bound homologous DNA at low-temperature conditions (as in Fig. 3).*

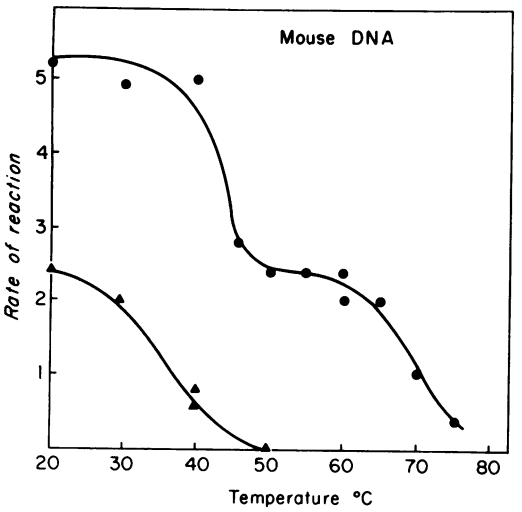

FIG. 9. *Rate of DNA duplex formation by labeled mouse DNA with homologous and heterologous DNA as a function of temperature. Conditions as in Fig. 2.* (●) *Mouse DNA.* (△) *Bacillus subtilis DNA.*

proportion depends on the chosen reaction conditions. It should be noted that none of these complexes has the stability expected for native or renatured DNA, i.e., T_m = 85 C. These results are analogous to those reported by Walker and McClaren (53), Martin and Hoyer (27), and Britten and Kohne (6), obtained from studies of strand association by the DNA-agar or hydroxy apatite methods.

Partially Redundant Nucleotide Sequences

The experiments reported with viral and bacterial DNA are consistent with the absence of closely related nucleotide sequences distributed around the genome. With T4 DNA, only two kinds of reaction product are formed, one with a stability close to that of native DNA and the other with a stability no higher than that of duplexes formed between two strands of unrelated DNA molecules. Thus, one may conclude that no two nucleotide sequences exist which have homology greater than that expected on a random basis. The situation with bacterial DNA is similar except for the evidence for some distant homologies significantly greater than those existing between two DNA preparations of the same base composition chosen at random.

In contrast, the experiments reported with mouse DNA show that the predominant reaction product formed is one between strands which are only partially complementary, representing different regions of the genome which are partially,

TABLE 2. *Nucleotide content of various genomes*

Organism	Mol wt (daltons)	Nucleotide pairs	Reference
Viruses			
φX174	0.9×10^6	5×10^{3a}	Sinsheimer (46)
T5 coliphage	84×10^6	130×10^3	Hershey et al. (17)
T2 coliphage	130×10^6	200×10^3	Rubenstein et al. (44)
Bacteria			
Mycoplasma gallisepticum	0.2×10^9	0.3×10^6	Morowitz et al. (36)
Haemophilus influenzae	0.7×10^9	1×10^6	Berns and Thomas (5)
Escherichia coli	2.6×10^9	4×10^6	Cairns (7)
Pseudomonas sp.	2.4×10^9	4×10^6	Park and DeLey (40)
Bacillus subtilis	2×10^9	3×10^6	Dennis and Wake (9)
	1.3×10^9	2×10^6	Massie and Zimm (28)
Yeast			
Saccharomyces cerevisiae	13×10^9	20×10^6	Ogur et al. (39)
Invertebrates			
Drosophila melanogaster[b]	0.12×10^{12}	0.2×10^9	Ritossa and Spiegelman (43)
Sea urchin (*Lytechinus*) sperm	0.6×10^{12}	0.8×10^9	Mirsky and Ris (34)
Vertebrates			
Chicken[b]	1.3×10^{12}	2.1×10^9	Mirsky and Ris (34)
Mouse[b]	3×10^{12}	4.7×10^9	Vendrely and Vendrely (52)
Human[b]	3×10^{12}	5.6×10^9	Vendrely and Vendrely (52)

[a] For double-stranded replicative form.
[b] Diploid cell; all others haploid.

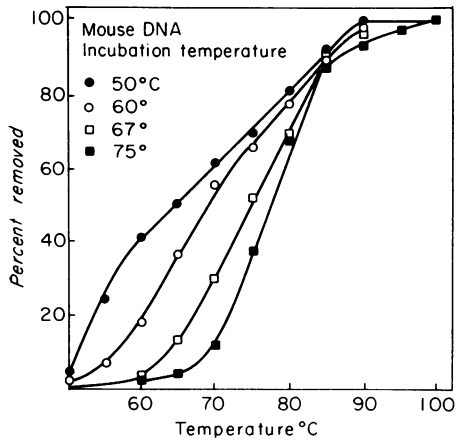

FIG. 10. *Thermal dissociation profiles of duplexes formed by incubating labeled mouse DNA with filter-bound homologous DNA at various temperatures.*

but not completely, similar in base sequence. Moreover, it is clear that many degrees of similarity of base sequence exist among many parts of the mouse genome, so that duplexes of various degrees of mismatching may be formed. This leads to the concept of families of related base sequences in which the individual members are related to one another to various degrees (6). It is important to note that not all of the base sequences in mouse DNA are partially redundant in this sense. Britten and Kohne (6) have shown that about half of the base sequences are unique and not members of families of related genes. Thus, if the concentration is high enough and the incubation time long enough, this fraction of the DNA may be completely renatured.

Thus, it is apparent that the interpretation of studies of DNA duplex formation or DNA-RNA hybrid formation requires some caution. For example, the reactions among primate DNA preparations (Table 1) are not measures of base sequence differences in individual genes of the various species but of relationships among families of base sequences. The reactions are not completely locus-specific, although as demonstrated by these results they do show considerable species specificity.

Correspondingly, although DNA-RNA hybrid formation is generally thought of as the reconstruction of the association between a gene and its primary gene product, this concept is justifiable only in viral and bacterial systems (8, 30). In mammalian systems, the reaction is one between an RNA molecule and a DNA molecule similar in base sequence to the DNA site which produced that RNA molecule. Results similar to those of Fig. 10 are obtained if hybrids are formed at different temperatures and then dissociated (8). This does not mean, of course, that hybridization reactions are not useful in the analysis of RNA synthesis in mammalian systems, but merely that the results are easily over-interpreted. For example, the degree of difference demonstrable between the population of RNA molecules in two tissues is highly dependent on the reaction conditions. The more stringent the conditions, the more differences appear. Thus, mouse liver and mouse kidney RNA appear quite similar in competition reactions carried out at 60 C and almost completely different when assayed at 72 C (Church and McCarthy, *in preparation*). This is easily interpretable, by comparison with the data of Fig. 10, as resulting from different degrees of cross-reaction between the RNA molecules and related DNA sequences. This means that the differences demonstrated between two populations of RNA molecules of mammalian origin are always minimal estimates of the actual differences, since the existence of related sequences limits resolution. In the same way, titration or saturation experiments, in which a determination of the maximal proportion of DNA sites which will react is made (55), are often misinterpreted. Since RNA molecules will react with sites other than those responsible for their synthesis, this assay will provide an overestimate of the fraction of the DNA sites actively synthesizing RNA in the situation in question. Notwithstanding these qualifications, the analysis of RNA by hybridization methods provides the most sensitive means presently at hand for comparison of populations of RNA molecules. Even in the limiting case of no locus specificity in the reaction of RNA with DNA, the method may be considered as a form of chromatography in which the DNA column contains a vast variety of different sites, each with its own affinity for a given group of RNA molecules. From this point of view it is apparent that differences among RNA populations may often be demonstrable, although the failure to detect differences does not establish identity.

CONSERVATISM IN THE EVOLUTION OF BASE SEQUENCES

The evolution of base sequences takes place at different rates at different sites in the genome. This is evident from comparative studies of individual proteins in various groups of organisms. In birds, for example, certain structural proteins evolve more rapidly than certain enzymes (45, 54). From more direct experiments with DNA, it has been shown that certain nucleotide sequences retain some homology between organisms as distant as fish and primates, while the majority of the genome has undergone much greater

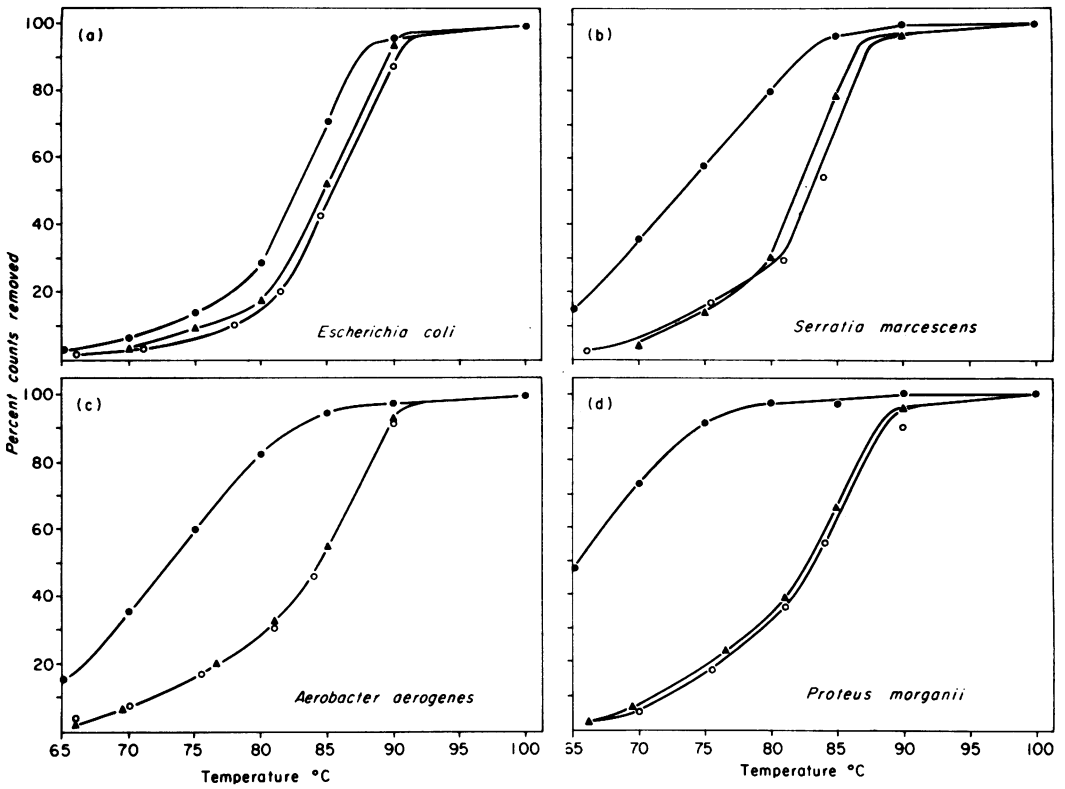

FIG. 11. *Comparison of the thermal dissociation profiles of hybrids formed with either Escherichia coli rRNA or E. coli pulse-labeled RNA. E. coli* ^{32}*P-labeled 23S or 16S RNA (10 μg; 9.4 × 10^5 counts per min per μg) was incubated with filters containing 50 μg of the various DNA preparations. The profiles for* ^{14}*C (pulse)-labeled RNA (179 counts per min per μg) were obtained by incubating 33.3 μg of RNA with filters containing 50 μg of DNA from various sources. Pulse-labeled RNA (●); 23S RNA (▲); 16S RNA (○). (a) Escherichia coli DNA filters; (b) Serratia marcescens; (c) Aerobacter aerogenes; (d) Proteus morgani. From Moore and McCarthy (35).*

change in base sequence (19). In addition, a core of relatively conserved sequences exists among all mammals (19).

Ribosomal RNA (rRNA) and sRNA cistrons are examples of conservative regions of the genome. These regions are represented in the DNA of many genomes by some 0.3 and 0.02% of the total, respectively (55, 56). Interspecific hybridization between *E. coli* sRNA and other enterobacterial DNA molecules occurs to a much greater extent than corresponding hybridization of pulse-labeled or mRNA (16). In *Bacillus*, enterobacteria, and myxobacteria, base sequences in rRNA are conservative relative to the total (11, 13, 35, 50). This is exemplified by the data of Fig. 11, showing a comparison between the thermal stability profiles of hybrids with DNA formed by *E. coli* pulse-labeled 16*S* and 23*S* rRNA (35). In each case, heterologous hybrids formed by rRNA show smaller differences in stability, compared with homologous hybrids, than is the case for pulse-labeled RNA (35). Similar results were obtained with *Myxococcus xanthus* RNA. Since, in this case, the base composition of the mRNA is richer in GC than is rRNA, the higher cross-reaction of rRNA is more readily attributable to conservatism in base sequence in these cistrons than to its intrinsic base composition (35).

Comparisons of the hybridization properties of rRNA of mammals and bacteria provide results quite similar to those obtained from studies of DNA duplex formation described above. Thus, the T_m of *E. coli* rRNA-DNA hybrids is close to that expected from its base composition (Moore and McCarthy, *in preparation*). The same thermal stability is obtained if the hybrids are treated with ribonuclease to remove grossly unpaired regions (15). The thermal stability of the hybrid is not dependent on the temperature of incubation (Fig. 12). On the other hand, the thermal stability of rabbit rRNA-DNA hybrids is lower, even though the base composition is higher in GC and the stability of the remaining complex is increased

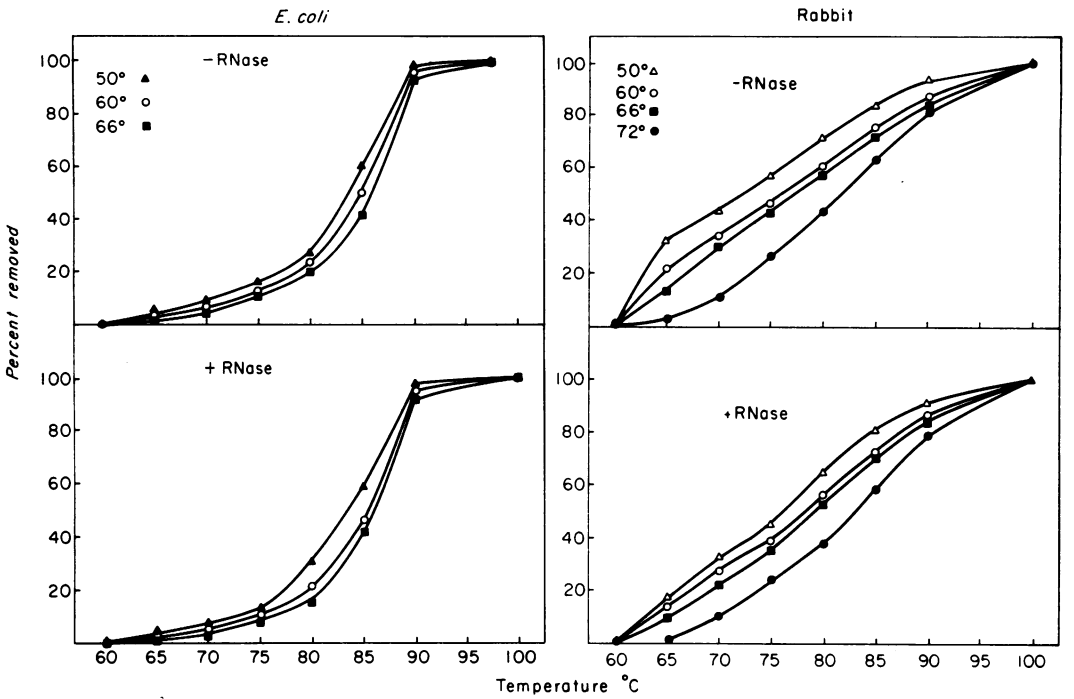

FIG. 12. *Thermal dissociation profiles of DNA-RNA hybrids formed by Escherichia coli 23S ribosomal RNA or rabbit 28S ribosomal RNA with homologous DNA by incubating at various temperatures. Profiles were determined with or without treatment with ribonuclease to eliminate some mispaired structures.*

after ribonuclease treatment. Moreover, as with DNA-DNA duplexes of mammalian origin, the properties of the hybrid depend on the reaction conditions employed during its formation (Fig. 12). These facts suggest that the many rRNA cistrons (about 500) in mammalian DNA are similar, but not identical, in sequence, so that hybridization occurs between rRNA having one base sequence and a DNA base sequence representing another rRNA species, a situation entirely analogous to that observed with DNA-DNA duplex formation.

INTERACTION OF OLIGONUCLEOTIDES WITH DNA

Specificity of Interaction

One of the obvious questions which arises from these experiments is: how does one correlate differences in the stability of duplexes with the numbers of complementary and noncomplementary bases in them? As mentioned at the outset, our aim is to define differences and similarities in base sequence between two genes of one organism or between corresponding genes in related organisms. This would furnish valuable information regarding the rate of evolution of base sequences in various groups of organisms and in various parts of the genome.

One approach involves the study of the interaction of fragments of DNA or oligodeoxyribonucleotides with DNA. Chemical methods of degradation and fractionation allow the preparation of oligonucleotides of known chain length representing all the base sequences of the parental DNA of that length (32). These may be used in reactions with single-stranded DNA to determine the stability and specificity of such interactions.

An example of this approach is given in Fig. 13. A mixture of mouse and *E. coli* oligonucleotides of mean chain length, $n = 33$, were incubated with both single-stranded parental DNA preparations. The complexes were then subjected to increasing temperature to dissociate them in the usual way. As is clear from the data, some specificity exists in the amount of complex formed (as represented by the 25 C points), and this increases as less well-paired structures are dissociated; at high temperatures, only species-specific complexes remain stable. Thus, in this size range, it is quite possible to demonstrate specificity.

Another experiment demonstrated that small fragments (of various sizes) of T4 phage DNA react nonspecifically and that specificity increases very rapidly with size (Fig. 14). Oligonucleotides derived from phage T4 DNA, in the size range

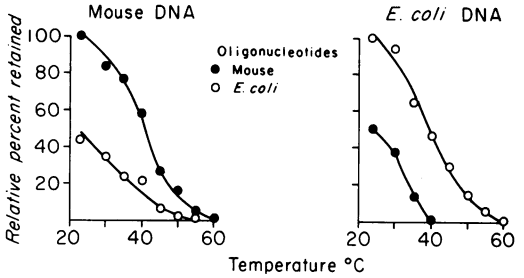

FIG. 13. *Thermal dissociation profiles of complexes formed by slow cooling 0.2 μg of a mixture of ^{14}C-labeled mouse oligodeoxynucleotides and ^{32}P-labeled Escherichia coli oligodeoxynucleotides of chain length n = 33 with 100 μg of mouse or E. coli DNA. From McConaughy and McCarthy (32).*

n = 11 to 28, were incubated with parental DNA and two types of unrelated DNA (32), and melting profiles of the complexes formed are shown. With the lowest size shown, n = 14, and below, all the points superimpose on one line, indicating a lack of specificity. At n = 17 and above, specific interactions dominate the reaction, and this specificity increases with increasing chain length. These results show that, as would be expected from elementary statistical considerations (32), oligonucleotides of chain length greater than about 14 to 15 show high specificity. Similar results from studies of hybrids between oligonucleotides of the ribose series and DNA have been reported (38*a*). This is clearly due to the failure to find highly complementary base sequences in heterologous DNA molecules. Interactions do, of course, occur with any DNA, but these do not occupy all of the bases contained in the chain. It is clear from these considerations that, with high-molecular-weight DNA, at least this extent of base-pairing between the two strands must occur for the interaction to be a specific one.

Estimation of the Degree of Mismatching of Bases

The same experiment may be used to provide information as to the number of base sequence similarities which occur between similar sites in the genome or corresponding sites in different DNA molecules. The principle is based on the fact that, in the size range, n = 3 to n = about 50, the stability of the duplex depends upon its length. Thus, the T_m increases with increasing chain length. This has been shown for many model systems of synthetic polynucleotides (23, 38, 42), and for the interactions of oligonucleotides and DNA (32).

Figure 15 illustrates a preliminary experiment aimed at using this principle to determine the extent of mismatching in the duplexes formed by mammalian DNA. The data are taken from a double-label experiment, of the type shown earlier, in which a differentially labeled mixture of oligonucleotides from the two species are incubated with both parental DNA preparations. The subtraction plot, in which the percentage reaction of heterologous oligonucleotides is subtracted from the homologous reaction, is a measure of the species-specific complexes. Even though the base compositions are the same, *B. subtilis* complexes are considerably more stable, indicating a lack of fidelity in matching for the mouse DNA. The displacement to lower temperatures is a measure of the degree of mismatching. Current experiments have shown that *B. subtilis* fragments of size n = 20 give a curve close to that for mouse DNA at size 25. We may, therefore, draw the tentative conclusion that about five base pairs are mismatched. This effect must be due to the arrangement of base sequences in mouse DNA itself and would be expected from the data derived from interactions of high-molecular-weight DNA. Because of the fact that there are many sites in DNA of related, but not identical, base se-

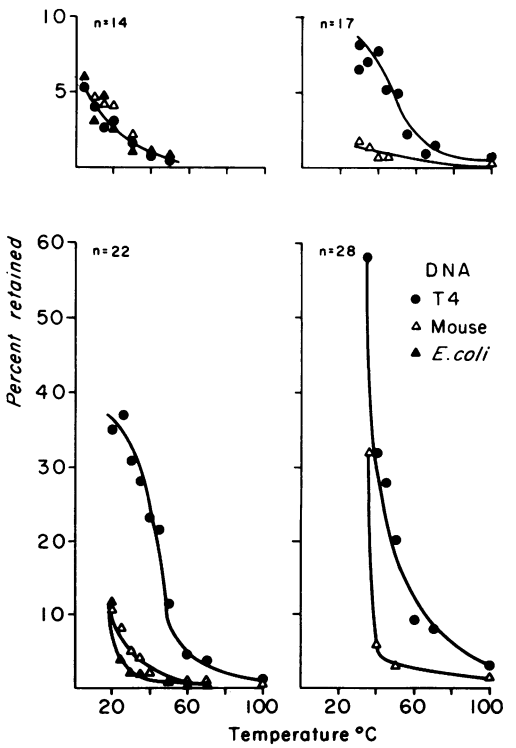

FIG. 14. *Melting profiles of complexes formed by incubating 8 μg of ^{32}P-labeled T4 phage oligodeoxynucleotides of various chain lengths with 100 μg of T4 phage, Escherichia coli, or mouse DNA. From McConaughy and McCarthy (32).*

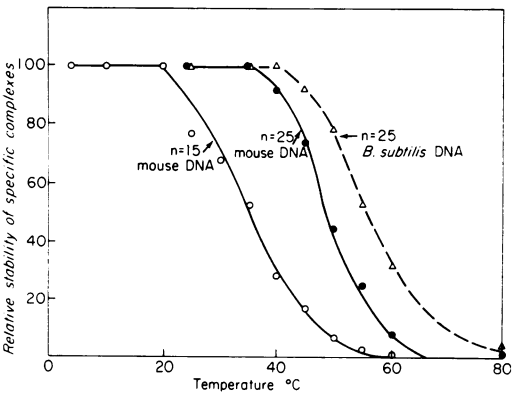

FIG. 15. *Relative stability of complexes formed by a mixture of oligodeoxyribonucleotides originating from mouse or Bacillus subtilis DNA (with n = 15 or 25) and the corresponding single-stranded DNA. The data represent the stability of species-specific complexes, obtained by subtracting the reaction of heterologous oligonucleotides from that of homologous oligonucleotides. An equimolar mixture of ^{3}H-labeled mouse and ^{32}P-labeled B. subtilis oligonucleotides (0.07 μg) was slowly cooled with 12 μg of filter-bound DNA.*

quence, these somewhat mispaired structures will dominate the spectrum of reaction products.

It should be emphasized that this estimate for the extent of mismatching is very crude and preliminary. It assumes that mismatched bases have no influence on the stability of the interaction complex. This is almost certainly an oversimplification. Nevertheless, it is clear that this approach can be highly rewarding and can give us quantitative information. Artificial in vitro modification of the bases to produce known degrees of mismatching can also be used to calibrate the method, to permit quantitative estimates of base sequence divergence.

The possibility of estimating base sequence divergence may be illustrated by a consideration of interspecific reactions. Figure 16 shows the interaction of a mixture of oligonucleotides, of $n = 25$ or $n = 40$, derived from mouse and *B. subtilis* DNA with various rodent DNA preparations. It is possible to determine whether each DNA is able to recognize the mouse oligonucleotides as those of the same or a related organism. Differentially labeled oligonucleotides originating from *B. subtilis* DNA are included as a control for nonspecific interactions. Thus, when the curves superimpose, no discrimination is made. Obviously, specific reactions are given with both sizes by parental DNA. On the other hand, base sequence complementarity offered by fragments of mouse DNA of $n = 25$ is not sufficient to give specific interactions with rat or hamster DNA. When n is increased to 40, the information content is sufficiently high to allow specific reactions with rat DNA. Even at this size, however, sufficient similarity in sequence does not exist for specific interactions to occur with hamster DNA. This DNA is related in base sequence, as demonstrated by reactions between high-molecular-weight mouse and hamster DNA (20) and by taxonomic considerations. Clearly, we can make estimates of the degree of pairing and therefore the extent of base sequence similarity between mouse and rat DNA by subtraction plots as shown earlier. These experiments could be carried out among any group of related DNA types to provide information as to extents of sequence divergence among species or within the DNA of one organism.

EVOLUTION AND RECOMBINATION OF DNA MOLECULES

During the evolution of life on earth, two kinds of changes in the DNA have accompanied the development of more complex forms. One involves changes in base sequence and the other a progressive increase in the DNA content of a genome. The former changes are developed by mutation leading to base substitution, or insertion and deletion of bases leading to a more radical change in base sequence. This diversification of base sequence is most easily seen in bacteria in which the overall base composition varies between 25 and 75% GC. This variation in base

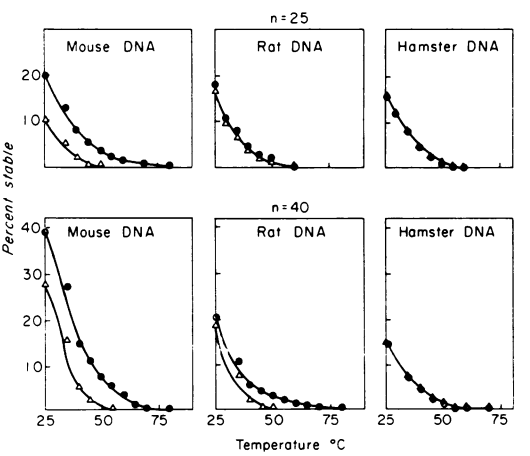

FIG. 16. *Thermal dissociation profiles of complexes formed by slow cooling a mixture of mouse and Bacillus subtilis oligodeoxyribonucleotides with mouse, rat, or hamster DNA bound to filters; 0.07 μg of an equimolar mixture of oligonucleotides (n = 25 or 40) of mouse and B. subtilis was slowly cooled with 12 μg of filter-bound DNA. (●) ^{3}H-labeled mouse oligonucleotides; (△) ^{32}P-labeled B. subtilis DNA.*

composition within a group also exists in protozoa, the more primitive metazoans, and higher plants, although the overall base composition of the DNA of vertebrates is quite similar. In all these cases, however, a considerable variation in base composition of segments within a genome is apparent.

As a general rule, the DNA content per haploid cell has increased progressively through the evolution of more complex forms. Viruses, with the smallest number of gene functions, contain the least DNA, and this quantity increases progressively through bacteria, protozoa, and metazoans to the vertebrates (Table 2). There are, however, some exceptions to this generalization, the meaning of which remains obscure. The most well-known example is that of the Anuran amphibians, whose DNA complement is many times that of other amphibia (29, 34). Other well-known examples have been documented in various families of higher plants (47). Whatever the meaning of these anomalies, it is reasonable to conclude that the predominant pattern is one of a steady increase in the DNA content of a genome during evolution, closely correlated with the structural and the behavorial complexity of the organism.

There are several possibilities that might be suggested for the mechanism by which more DNA is created to specify newly evolving functions. Such new base sequences may possibly be generated de novo by end addition to existing DNA molecules or by reiterative synthesis (22). Such base sequences would presumably be random, although useful sequences might conceivably evolve during an extended period. Alternatively, an increase in the total DNA might be accomplished through the acquisition of ready-made sequences from within the genome or other external sources. In bacteria, transduction and sexduction lead to an increase in the amount of DNA. The latter process, in particular, is known to operate across species boundaries and would therefore allow the acquisition of genetic material from different species. After sexduction or transduction, the integrated episomal genes may become a stable part of the bacterial chromosome. It is quite possible that similar mechanisms may operate in higher organisms, since clear evidence exists that the genome of tumor viruses may enter into a stable association with the host cell nucleus (3).

Notwithstanding these other possibilities, it seems most likely that gene duplication followed by translocation is a major contributor to the increasing DNA content in the genome. Much evidence for this process derives from determinations of amino acid sequence in similar proteins. A particularly persuasive example is provided by the various hemoglobin chains (58). It is evident from the studies reported here and those of other workers (6) that this mechanism has been an important one during the evolutionary history of mammalian DNA.

In a very simple DNA, as represented by T4 phage DNA, there is no evidence for the existence of duplications. Thus, it may be that this process does not occur or is rapidly obscured by divergence, or that the generation of such duplicated base sequences is for some reason disadvantageous. In this regard, the arguments set forth by Thomas (51) would tend to rationalize this finding. If, as seems most likely, a basic stage in the recombination of DNA molecules involved the association of complementary base sequences (33) between the two DNA molecules, the fidelity of recombination would be disturbed by the existence of very similar base sequences around the chromosome. Since the unequal crossing-over generated by this effect would, in most cases, be lethal (51), the persistence of gene duplications would be unlikely unless the duplicated genes can rapidly be made dissimilar by mutation and selection.

Similar conclusions may be drawn from the experiments with *B. subtilis* DNA. No closely similar base sequences exist, although there is evidence for very distant intragenome homologies. The rate of the duplex-forming reaction under these conditions suggests that base sequence similarities at this level are shared by many regions of the DNA. This may well be the result of historically remote gene duplications. Freese and Yoshida (14) presented evidence for homology among three *B. subtilis* dehydrogenases, suggesting that these have evolved from a common DNA region during bacterial evolution. Other explanations may be offered; for example, it may be that certain sequences of codons occur quite frequently as a result of the existence of similar active sites in many enzymes. In either case, apart from the rRNA cistrons, it seems clear that base sequence relationships among different cistrons above a certain level do not exist. Again, it is possible that they are, in fact, precluded by the requirements of recombination events.

In mouse DNA, the situation is quite different, so that a very complex set of related base sequences exist. If the existence of similar base sequences along the chromosomes does indeed present a problem for recognition of events preceding recombination, it seems most likely that this is surmounted by the compartmentalization of DNA among chromosomes. The synapsis of chromosomes would serve to restrict the spatial distribution of potentially recombining DNA

base sequences so that strand recombination could take place only between allelic segments.

At the moment, the relationship between the chromosome complement and the distribution of base sequences in the DNA is far from clear, since few DNA preparations have been studied in detail. The question as to whether related base sequences are randomly distributed among chromosomes cannot be answered until individual chromosomes may be prepared. Nor is it known whether there is any correlation between eucaryotic chromosome structure and the presence of related base sequences in the DNA. Whether the discontinuity between procaryotic and eucaryotic cells is also represented as a boundary between simple and complex partially redundant DNA molecules remains to be seen.

ACKNOWLEDGMENTS

This research was supported by grants from the National Science Foundation (GB 6099) and from the National Institutes of Health, Public Health Service (GM 12449). This research would not have been possible without the eager cooperation of my students and colleagues, R. B. Church, C. D. Laird, B. L. McConaughy, and R. L. Moore.

LITERATURE CITED

1. BAUTZ, E. K. F., AND E. REILLY. 1966. Gene-specific messenger RNA: isolation by the deletion method. Science **151**:328–330.
2. BELLETT, A. J. D. 1967. Numerical classification of some viruses, bacteria and animals according to nearest-neighbor base sequence frequency. J. Mol. Biol. **27**:107–112.
3. BENJAMIN, T. L. 1966. Virus specific RNA in cells productively infected or transformed by polyoma virus. J. Mol. Biol. **16**:359–373.
4. BERG, P., H. FANCHER, AND M. CHAMBERLIN. 1963. The synthesis of mixed polynucleotides containing ribo- and deoxyribonucleotides by purified preparations of DNA polymerase from *E. coli*, p. 467–483. *In* V. Bryson and H. J. Vogel [ed.], Informational macromolecules. Rutgers Symposium. Academic Press, Inc., New York.
5. BERNS, K. I., AND C. A. THOMAS, JR. 1965. Isolation of high molecular weight DNA from *Hemophilus influenzae*. J. Mol. Biol. **11**:476–490.
6. BRITTEN, R. J., AND D. E. KOHNE. 1966. Nucleotide sequence repetition in DNA. Ann. Rept. Carnegie Inst. Wash. **65**:78–106.
7. CAIRNS, J. 1963. The chromosome of *E. coli*. Cold Spring Harbor Symp. Quant. Biol. **28**:43–45.
8. DENIS, H. 1966. Gene expression in amphibian development. I. Validity of the method used: interspecific and intraspecific hybridization between nucleic acids. Properties of mRNA synthesized by developing embryos. J. Mol. Biol. **22**:269–283.
9. DENNIS, E. S., AND R. G. WAKE. 1966. Autoradiography of the *Bacillus subtilis* chromosome J. Mol. Biol. **15**:435–439.
10. DENHARDT, D. T. 1966. A membrane-filter technique for the detection of complementary DNA. Biochem. Biophys. Res. Commun. **23**:641–646.
11. DOI, R. H., AND R. T. IGARASHI. 1965. Conservation of ribosomal and messenger ribonucleic acid cistrons in *Bacillus* species. J. Bacteriol. **90**:384–390.
12. DOTY, P., J. MARMUR, J. EIGNER, AND C. SCHILDKRAUT. 1960. Strand separation and specific recombination in DNA's. II. Physical chemical studies. Proc. Natl. Acad. Sci. U.S. **46**:461–476.
13. DUBNAU, D., I. SMITH, P. MORELL, AND J. MARMUR. 1965. Gene conservation in Bacillus species. I. Conserved genetic and nucleic acid base sequence homologies. Proc. Natl. Acad. Sci. U.S. **54**:491–498.
14. FREESE, E., AND A. YOSHIDA. 1965. The role of mutations in evolution, p. 341–355. *In* V. Bryson and H. J. Vogel [ed.], Evolving genes and proteins. Rutgers Symposium. Academic Press, Inc., New York.
15. GILLESPIE, D., AND S. SPIEGELMAN. 1965. A quantitative assay for DNA/RNA hybrids with DNA immobilized on a membrane. J. Mol. Biol. **12**:829–842.
16. GOODMAN, H. M., AND A. RICH. 1962. Formation of a DNA soluble RNA hybrid and its relation to the origin, evolution and degeneracy of soluble RNA. Proc. Natl. Acad. Sci. U.S. **48**: 2101–2109.
17. HERSHEY, A. D., E. BURGI, AND L. INGRAHAM. 1962. Sedimentation coefficient and fragility under hydrodynamic shear as measures of molecular weight of the DNA of phage T5. Biophys. J. **2**:423–431.
18. HOLLEY, R. W., J. APGAR, G. A. EVERETT, J. T. MADISON, M. MARQUISEE, S. H. MERRILL, J. R. PERSWICK, AND A. ZAMIR. 1965. Structure of a ribonucleic acid. Science **147**:1462–1465.
19. HOYER, B. H., B. J. MCCARTHY, AND E. T. BOLTON. 1964. A molecular approach in the systematics of higher organisms. Science **144**:959–967.
20. HOYER, B. H., E. T. BOLTON, B. J. MCCARTHY, AND R. B. ROBERTS. 1965. The evolution of polynucleotides, p. 581–590. *In* V. Bryson and H. J. Vogel [ed.], Evolving genes and proteins Rutgers Symposium. Academic Press, Inc., New York.
21. JOSSE, J., A. D. KAISER, AND A. KORNBERG. 1961. Enzymatic and synthesis of DNA. VIII. Frequencies of nearest-neighbor base sequences in DNA. J. Biol. Chem. **236**:864–871.
22. KORNBERG, A. 1965. Synthesis of DNA-like polymers *de novo* or by reiterative replication, p. 403–417. *In* V. Bryson and H. J. Vogel [ed.], Evolving genes and proteins. Rutgers Symposium. Academic Press Inc., New York.
23. LIPSETT, M. N. 1964. Complex formation between polycytidylic acid and guanine oligonucleotides. J. Biol. Chem. **239**:1256–1260.
24. MARMUR, J., AND D. LANE. 1960. Strands separation and specific recombination in DNA's. I. Biological studies. Proc. Natl. Acad. Sci. U.S. **46**:453–461.

25. MARMUR, J., AND P. DOTY. 1961. Thermal renaturation of DNA. J. Mol. Biol. **3**:585–594.
26. MARMUR, J., R. ROWND, AND C. L. SCHILDKRAUT. 1963. Denaturation and renaturation of DNA. Progr. Nucleic Acid Res. **1**:231–300.
27. MARTIN, M. A., AND B. H. HOYER. 1966. Thermal stabilities and species specificities of reannealed animal DNA's. Biochemistry **5**:2706–2713.
28. MASSIE, H. R., AND B. H. ZIMM. 1965. Molecular weight of the DNA in the chromosomes of *E. coli* and *B. subtilis*. Proc. Natl. Acad. Sci. U.S. **54**:1636–1641.
29. MCCARTHY, B. J. 1965. The evolution of polynucleotides. Progr. Nucleic Acid Res. **4**:129–160.
30. MCCARTHY, B. J. 1967. General aspects of homology studies with nucleic acids. Intern. Symp. Tumor Viruses, 1st, Subviral Carcinogenesis, p. 43–61.
31. MCCARTHY, B. J., AND E. T. BOLTON. 1963. An approach to the measurement of genetic relatedness among organisms. Proc. Natl. Acad. Sci. U.S. **50**:156–164.
32. MCCONAUGHY, B. L., AND B. J. MCCARTHY. 1967. The interaction of oligodeoxynucleotides with denatured DNA. Biochim. Biophys. Acta **149**: 180–189.
33. MESELSON, M. 1964. On the mechanism of genetic recombination between DNA molecules. J. Mol. Biol. **9**:734–745.
34. MIRSKY, A. E., AND H. RIS. 1951. The DNA content of animal cells and its evolutionary significance. J. Gen. Physiol. **34**:451–462.
35. MOORE, R. L., AND B. J. MCCARTHY. 1967. Comparative study of ribosomal cistrons in enterobacteria and myxobacteria. J. Bacteriol. **94**: 1066–1074.
36. MOROWITZ, H. J., M. E. TOURTELLOTTE, W. R. GUILD, E. CASTRO, C. WOESE, AND R. C. CLEVERDON. 1962. The chemical composition and submicroscopic morphology of *Mycoplasma gallisepticum*. J. Mol. Biol. **4**:93–103.
37. MORRISON, J. M., H. M. KEIR, H. SUBAK-SHARPE, AND L. V. CRAWFORD. 1967. Nearest-neighbor base sequence analysis of the DNA of a further three mammalian viruses. Simian virus 40, human papilloma and adenovirus type 2. J. Gen. Virol. **1**:101–108.
38. NAYLOR, R., AND P. T. GILHAM. 1966. Studies on some interactions and reactions of oligonucleotides in aqueous solution. Biochemistry **5**:2722–2728.
38a. NIYOGI, S. K., AND C. A. THOMAS, JR. 1967. The specific association of ribo-oligonucleotides of known chain length with denatured DNA. Biochem. Biophys. Res. Commun. **26**:51–57.
39. OGUR, M., S. MINCKLER, G. LINDERGREN, AND C. C. LINDERGREN. 1952. The nucleic acids in a polyploid series of *Saccharomyces*. Arch. Biochem. Biophys. **40**:175–184.
40. PARK, I. W., AND J. DELEY. 1967. Ancestral remnants in DNA from *Pseudomonas* and *Xanthomonas*. Antonie van Leeuwenhoek J. Microbiol. Serol. **33**:1–16.
41. RAJBHANDARY, U. L., A. STUART, R. D. FAULKNER, S. H. CHANG, AND H. G. KHORANA. 1966. Nucleotide sequence studies on yeast phenylalanine sRNA. Cold Spring Harbor Sym. Quant. Biol. **31**:425–434.
42. RICH, A. 1960. A hybrid helix containing both deoxyribose and ribose polynucleotides and its relation to their transfer of information between the nucleic acids. Proc. Natl. Acad. Sci. U.S. **46**:1044–1053.
43. RITOSSA, F. M., AND S. SPIEGELMAN. 1965. Localization of DNA complementary to ribosomal RNA in the nucleolus organizer region of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. U.S. **53**:737–745.
44. RUBENSTEIN, I., C. A. THOMAS, JR., AND A. D. HERSHEY. 1961. The molecular weights of T2 bacteriophage DNA and its first and second breakage products. Proc. Natl. Acad. Sci. U.S. **47**:1113–1122.
45. SIBLEY, C. G. 1965. The characteristics of specific peptides from single proteins as data for classification, p. 435–450. *In* V. Bryson and H. J. Vogel [ed.], Evolving genes and proteins. Rutgers Symposium. Academic Press Inc., New York.
46. SINSHEIMER, R. L. 1959. A single-stranded DNA from bacteriophage ϕ X174. J. Mol. Biol. **1**: 43–53.
47. STEBBINS, G. L. 1966. Chromosomal variation and evolution. Science **152**:1463–1469.
48. SUBIRANA, J. A. 1966. Kinetics of renaturation of denatured DNA. II. Products of the reaction. Biopolymers **4**:189–200.
49. SUBIRANA, J. A., AND P. DOTY. 1966. Kinetics of renaturation of denatured DNA. I. Spectrophotometric results. Biopolymers **4**:171–187.
50. TAKAHASHI, M., M. SAITO, AND Y. IKEDA. 1967. Species specificity of the ribosomal RNA cistrons in bacteria. Biochim. Biophys. Acta **134**: 124–133.
51. THOMAS, C. A. JR., 1966. Recombination of DNA molecules. Progr. Nucleic Acids Res. Mol. Biol. **5**:315–348.
52. VENDRELY, R., AND C. VENDRELY. 1949. La teneur du noyau cellulaire en acide desoxyribonucléique à travers les organes, les individus et les espèces animales. Experientia **5**:327–329.
53. WALKER, P. M. B. AND A. MCCLAREN. 1965. Specific duplex formation *in vitro* of mammalian DNA. J. Mol. Biol. **12**:394–409.
54. WILSON, A. C., AND N. O. KAPLAN. 1965. Enzyme structure and its relation to taxonomy, p. 321–346. *In* V. Bryson and H. J. Vogel [ed.], Evolving genes and proteins. Rutgers Symposium. Academic Press Inc., New York.
55. YANKOFSKY, S., AND S. SPIEGELMAN. 1962. The identification of the ribosomal RNA cistron by sequence complementarity. II. Saturation of and competitive interaction at the rRNA cistron Proc. Natl. Acad. Sci. U.S. **48**:1466–1472.
56. YANKOFSKY, S., AND S. SPIEGELMAN. 1963. Dis-

tinct cistrons for the two ribosomal RNA components. Proc. Natl. Acad. Sci. U.S. **49:** 538–544.

57. ZACHAU, H. G., D. DUTTING, H. FELDMANN, F. MELCHERS, AND W. KARAU. 1966. Serine specific transfer RNA's. XIV. Comparison of nucleotide sequences and secondary structure models. Cold Spring Harbor Symp. Quant. Biol. **31**:417–424.
58. ZUCKERKANDL, E., AND L. PAULING. 1965. Evolutionary divergence and convergence in proteins, p. 97–166. *In* V. Bryson and H. J. Vogel [ed.], Evolving genes and proteins. Rutgers Symposium. Academic Press Inc., New York.